DOCUMENT RESUME

ED 178 879 CS 005 115

AUTHOR Snodgrass, Joan Gay
TITLE Toward a Model for Picture and Word Processing.
PUB DATE Sep 79
NOTE 36p.; Paper presented at the Annual International Conference on Processing of Visible Language (2nd, Niagara-on-the-Lake, Canada, September 3-7, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Association (Psychological); Associative Learning; Memory; *Models; *Pictorial Stimuli; *Reading Processes; Reading Research; *Recall (Psychological); Thought Processes; Visualization; Visual Stimuli; *Word Recognition

ABSTRACT

A model was developed to account for similarities and differences between picture and word processing in a variety of semantic and episodic memory tasks. The model contains three levels of processing: low-level processing of the physical characteristics of externally presented pictures and words; an intermediate level where the low-level processor makes contact with prototypical information about the looks of objects or pictures and the sounds of words; and the deepest (propositional) level where meaning is analyzed. The relationships between pictures and their names (or between visual and acoustic images) can take place directly, between the two image stores, or indirectly, via the propositional level to which both image stores have access. One difference between pictures and their names is the greater variability of appearances that objects and pictures have. A second difference is that pictures have less ambiguity of reference than their names, resulting in fewer propositional memory nodes accessed by pictures than by words. These differences are consistent with a large body of literature on picture-word processing differences. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178879

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Toward a model for picture and word processing

Joan Gay Snodgrass

New York University

A paper prepared for Processing of Visible Language 2

Niagara-on-the-Lake, Canada
3-7 September, 1979

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joan Gay Snodgrass

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Mailing address:
Joan Gay Snodgrass
Department of Psychology
New York University
6 Washington Place, Room 856
New York, New York 10003

C S005115

Abstract

A model is proposed to account for similarities and differences between picture and word processing in a variety of semantic and episodic memory tasks. The model contains three levels of processing, with the most superficial concerned with low-level processing of the physical characteristics of externally presented pictures and words; an intermediate level in which the results of the low-level processer make contact with prototypical information about how objects (or the pictures which represent them) look and how words sound; and the deepest (propositional) level in which meaning is analyzed. The interlingua between pictures and their names (or between their visual and acoustic images) can take place either directly, with connections between the two image stores, or indirectly, via the propositional level to which both image stores have access.

Two differences which emerge between pictures and their names are: (a) greater variability in the way objects or pictures appear compared to the way names appear or sound, which leads to greater variability in prototypical visual images than in prototypical acoustic images; and (b) less ambiguity of reference for pictures than for their names, with correspondingly fewer propositional memory nodes accessed by pictures than by words. These differences are shown to be consistent with a large body of literature on picture-word processing differences.

Toward a model for picture and word processing

Joan Gay Snodgrass

New York University

Processing of pictures and words has been compared in many different tasks and for many different reasons. In order to systematize this literature somewhat, I shall classify the tasks into two broad categories -- those studying semantic memory and those studying episodic memory. Semantic memory is assumed to be a general knowledge store representing a systematic and organized set of information about the world which is shared by members of the same cultural group. In contrast, episodic memory is autobiographical and idiosyncratic, and represents information about the occurrence of events in time and space for that individual (Tulving, 1972).

Examples of semantic memory tasks are naming pictures and words, generating images to words, categorizing pictures or their names, and making symbolic comparison judgments about whether X has more of some property than Y. In all of these tasks, subjects must base their judgments on knowledge of the world they bring with them to the laboratory. This knowledge includes information about how words are pronounced, by what name or names pictures are called, and whether a particular stimulus is a member of category X or Y.

Episodic memory tasks, on the other hand, require that subjects first learn a set of materials in the laboratory before being tested on it. When meaningful material such as pictures of known objects or their names are used, subjects are not, strictly speaking, learning about the items themselves, but rather learning that they occurred in a particular context at a particular time and place. Episodic memory tests include the standard verbal learning and memory paradigms such as free and serial recall, paired-associate recall, yes/no or forced-choice recognition, serial and spatial reconstruction and recognition, and so forth.

In order to relate this body of literature to the proposed model for picture-word processing, I will first present the model, next show how the model is related to other models in the literature, and finally relate the empirical data on picture-word processing differences to the proposed model.

The model was developed to account for processing differences between the type of pictures and words shown in Figure 1 -- that is, relatively

---

Insert Figure 1 about here

---

simple line drawings of common objects, and the names most commonly used to identify the drawings. Other possible pictorial-verbal pairs of items not considered are abstract paintings and their names; faces and their names; and complex scenes and their descriptions. I shall confine myself to this relatively simple set of stimuli primarily because most of the experimental data about picture-word processing differences are based upon such types of materials.

A model for picture and word processing

Figure 2 presents a schematic representation of the model. We will first

---

Insert Figure 2 about here

---

consider its operation from an external-to-internal (top-to-bottom) processing sequence.

External input may be in the form of either visually or auditorily presented words, and visually presented pictures. The first level of processing consists of the raw verbal or visual codes resulting from some relatively low-level processing of the external stimulus. Level I includes such physical characteristics as voice and intonation or typeface and color for auditorily and visually presented words, respectively; and orientation, amount of detail, and any peculiar characteristics not an integral part of an object's characteristics for pictures.

Level II consists of prototypical information about acoustic and visual images. Specifically, it contains information about how words sound, for the acoustic image store, and how objects (or the pictures that represent them) look, for the visual image store. Such information is acquired over the course of a person's experience with speaking and listening to language, and with seeing objects and pictures and, sometimes, of drawing them. The acoustic image prototype store corresponds to the products of inner speech, and is assumed to be the same store accessed in verbal thinking, while the visual image prototype store is assumed to be accessed in visual thinking or imaging.

The prototypical image stores are prototypical in the sense that they contain features which represent the basic characteristics of acoustic and visual images, but leave out nonessential or idiosyncratic detail which sometimes accompanies instances of words and objects or pictures. The represent the average way the word "apple" sounds or the average way the object "apple" looks.

Each image in the store can be in a potential or in a generated form. A potential image consists of a set of features, with a probability distribution across the possible values each feature can be assigned. Generating an image consists of picking a value for each feature and then using these values to generate the image. This generated image is then available to introspection. Thus a string of generated acoustic images represents "internal speech" and a set of generated visual images represents "visual or mental imagery." It seems likely that the acoustic and visual image generators are of limited capacity, perhaps limited to producing a single acoustic or visual image at a time, although the two types of generators may work in parallel. Thus, it may be possible to have simultaneous acoustic and visual images but not two simultaneous acoustic images or two simultaneous visual images. On the other hand, the potential image store is of unlimited capacity.

The features of the potential acoustic images might consist of lists of phonemes, number of syllables, sequencing rules, and the like along with pitch period, formant frequencies, and intonation, while the features for visual images might consist of shape, orientation, size, detail, color, and the like. Some features of the potential image might have a default character: for example, pitch period and formant frequencies similar in value to the speaker's own voice in the acoustic case, and a standard orientation and size in the visual case.

It seems plausible that there is more inter-and intra-subject consistency in the values assigned to features for acoustic images than for visual images (i.e., the probability distributions for each feature are less variable in the former than the latter). That is, there are more ways to image a concept than to name it. For example, subjects rate the degree of agreement between their internally generated name to a picture and the picture's name higher than between their internally generated image to a word and the word's picture (Snodgrass, Wasser, Finkelstein, & Goldberg, 1974).

Products of both Level I, physical analysis, and Level II, image generation, are available to introspection. In addition, it seems likely that subjects can also access information in potential image stores without necessarily generating the image. We assume this because of such phenomena as "tip-of-the-tongue" states, in which subjects seem to be aware of some, but not all, of the features of a potential acoustic image, and what we might call "tip-of-the-mind's eye" states, in which subjects cannot generate an image of a concept even though they know its meaning and some of its visual characteristics.

The potential image stores also serve as the pattern recognition systems. A pictorial or verbal pattern is recognized as being of a known picture, object, or word if enough of its features correspond to the feature lists in one of the image stores. When the potential image is used for recognition,

the recognition system acts like a feature-analysis system and probably operates in parallel, as in "Pandemonium" (see Neisser, 1967). When generated images are used, the recognition system acts like a template-matching system and, given the limited capacity of the image generation system, probably operates in serial. It seems likely that subjects may need to resort to generation for ambiguous or noisy pictorial and verbal stimuli.

In the process of recognizing speech or written language and of recognizing pictorial input, subjects also accumulate information about the degree of mismatch between Level I, the physical image store, and Level II, the prototypical image store. We assume that the mismatch processer will accumulate more mismatch information if a spoken word differs from the observer's internal voice, if a visually-presented word has a peculiar or unexpected typescript or color, or if a picture does not correspond closely to the prototypical image.

Level III is the propositional store, which is accessible by both acoustic and visual image prototypes. The propositional store is viewed as an abstract set of nodes and of interconnections among nodes which differ qualitatively and could be labeled by such relationships among nodes as "is a coordinate of", "is a member of the category", "has the property", and so on. Although we use language to talk about the nodes in propositional memory and their interrelationships, propositional memory has no language (or for that matter visual images) and its operations are not accessible to introspection (hence it is labeled as being at the unconscious level). Rather, relationships and nodes are connected to the acoustic image store on the one hand, and to the visual image store on the other, but in the latter case only for concepts and relationships which are picturable.

While particular visual and acoustic images can access common nodes in propositional memory, many acoustic images will access nodes not accessible by visual images. Thus, the two are shown as having only partial overlap, with words shown as accessing more propositional "space" than pictures. The nature of many tasks may limit the number of propositional nodes a word will access: for example, when subjects are told to form visual images of words in a memory experiment, they may only access nodes in propositional memory which are also connected to the visual image store. However, many words have several picturable senses, for example, "pipe", "nut", "bat", and "top". Thus, access to one node may lead to one visual image (e.g., baseball bat) while access to another may lead to a different image (e.g., flying rodent). Thus a single acoustic image may access more than one node in Level III.

Figure 2 shows direct connections between the acoustic and visual image prototype stores. This allows for the possibility that picture-name matches may be made without accessing the semantic memory store, although this direct matching mode seems not to be used very often, for reasons we consider in a moment.

So far, we have considered the operation of the schema from an top-down or external to internal way. The operation of it from a bottom-up or internal to external way is analogous. Thinking, for example, is the activation of propositional memory nodes which in turn activate either acoustic image prototypes (in verbal thinking), visual image prototypes (in visual thinking), or both. We have assumed that propositional memory is unconscious because it seems difficult, on the basis of introspection, to know what we will think next until we have thought it, what we will say next until we have said it, what we will dream next until we have dreamt it. On the other hand, input from propositional memory can be used to disambiguate "meaning" of thoughts.

For example, if I think "tomorrow I will buy a TOP", the acoustic image prototype that is activated is identical whether I mean a toy, a blouse, or a coffee pot lid, yet there is no ambiguity in my understanding of my thought. This might be accounted for either by the fact that the propositional node from which the acoustic image "TOP" is generated can be known to me, from the possibility that the same propositional node also activates the relevant potential or generated visual image of the toy, or by both. Thus, while the operation and activation of propositional memory is not available to introspection, the input nodes from which images are generated may be. This problem of potential ambiguity of verbal thoughts suggests that adults may use input from propositional memory to disambiguate the meaning of thoughts (either through direct access to the node from which input is coming or by awareness of accompanying visual imagery). Since propositional activation is useful in disambiguating thinking, it may be that in tasks where such activation is not necessary (such as picture-naming or picture-word matching) subjects continue to access the propositional store because in the past it has been useful.

Relationship of the proposed model to other memory models

All three levels of the proposed model depicted in Figure 2 may be considered parts of semantic memory, in the sense it was defined previously. Level I must include knowledge about how to disambiguate and categorize the products of perception, Level II must include knowledge about the appearance of words and things, and Level III must include knowledge about types of associations among concepts.

Level III, the propositional store, most closely resembles the concept of semantic memory as it has been used in the literature, and the acoustic image store of level II most closely resembles the concept of the internal lexicon. The lexicon is usually assumed to

include information about graphemic as well as phonemic and acoustic features of words. In the present model, the graphemic part has been omitted for simplicity, although a complete model clearly requires this component. Emphasis has been placed upon the acoustic image because of the large amount of evidence that phonemic encoding of visually-presented words occurs at least some of the time. As Nelson (1979) has put it, "As far as words are concerned, hearing what is seen appears to be more likely than seeing what is heard." In any case, the model does not require that an acoustic image be generated during visual word recognition (a step implied by the phonemic recoding hypothesis), only that its potential image code be accessed.

Both the acoustic and visual image stores of Level II resemble feature analysis models of pattern perception in their potential form, and template-matching models in their generated form. Finally, Level I analysis includes iconic and echoic perceptual processes, and the residue of the pattern recognition process accumulated in the mismatch processer.

An important aspect of the model is its multipurpose use of stores and their processes for analyzing incoming information, carrying on internal processes such as thinking, and generating external responses. So, for example, the visual image store can be used to disambiguate visual pictures and scenes, to generate visual images during thinking, and to generate a visual image to a spoken word for comparison with a picture.

Because aspects of the model embrace almost all of cognitive psychology, I will concentrate on those aspects of the model which are of particular importance in picture-word processing differences: namely, the operation of the two image stores and their relation both to each other and to the propositional store. Within this limited domain, some crucial issues to be considered are: (a) the nature of the propositional store, including its unconscious status; (b) the level of abstractness at which concepts are represented in the visual image store;

(c) whether a given surface form of a word or picture (or its corresponding image) has multiple representations in the propositional store; (d) the relationship of the model to dual-code versus propositional models; and (e) the relationship between episodic and semantic memory. To avoid redundancy, the last issue will be discussed in the section on episodic memory tasks.

Theories of semantic memory: As Smith (1978) pointed out in his recent review of semantic memory models, models of semantic memory can range from very broad views which include both "semantic" and "episodic" memory events in the sense described previously (e.g., Anderson & Bower, 1973; Kintsch, 1974; Norman & Rumelhart, 1975); somewhat narrower views encompassing only information in what we have called semantic memory, but including other forms of knowledge than linguistic (such as the present model); and finally, the most restricted set which considers only words, their meanings, and the rules for operating on them.

Most models of semantic memory which label themselves as such fall into this latter category. These models correspond roughly to propositional memory and its connections to the acoustic image store in the present model. Within this restricted domain, theories of semantic memory can be classified into network models (e.g., Quillian, 1967; Collins & Quillian, 1972a, b; Collins & Loftus, 1975; Wickelgren, 1979) and feature-comparison models (Rips, Shoben & Smith, 1973; Smith, Shoben & Rips, 1974).

The propositional component of the proposed model most closely resembles a network model with labeled associations, although the visually and acoustically representable properties of both nodes and associations are retrieved from the visual and acoustic image stores. The meaning of a concept is defined as the sum of a node's associations with other nodes in the network, although the present model does not propose a particular structure for the network (in particular, no hierarchical structure is assumed).

A basic problem in the present model is to decide at what level of abstractness visual images are stored. At the most general level, one could imagine that the visual image store consisted of a single vector the selection of whose values could generate any possible visual image. At the most specific level, one could imagine a vast array of entries corresponding to each and every possible appearance of each and every object. I assume in the model that images are stored at an intermediate level of abstractness, corresponding to what semantic memory theorists identify as exemplars of categories (e.g., dog, canary, chicken, tomato). However, these exemplars in turn can be considered categories for such subordinate concepts as collie, pullet, plum tomato, etc. Rosch and her colleagues (Rosch, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Rosch, 1978) have proposed that a basic conceptual level for most tasks occurs at a level of categorization called the basic level, defined by a variety of criteria, including similarity of appearance, function, and motor involvement. In accord with this principle, the model assumes that the visual image store consists of separate entries for basic level categories, and that subordinate concepts to the basic level are generated by selecting particular (and unusual) values for the features. For example, the image of a plum tomato would be generated by selecting the less probable value on the shape feature of oblong rather than round.

Perhaps the most straightforward criterion for whether a concept is at the basic-level in the context of the present model is that it can be drawn so that most subjects identify it by the same name (Snodgrass & Vanderwart, 1979). This would mean that they had a separate entry for it in their image stores. Snodgrass and Vanderwart (1979) found, in agreement with Rosch et al. (1976), that basic level categories include some which are usually considered to be superordinates, such as bird, tree, fish and flower, and some which are usually considered to be atypical subordinates, such as chicken, duck, ostrich and penguin. Thus while most semantic memory theorists

look to formal class inclusion relationships to define exemplars and close and distant superordinates, the present model relies on pragmatic considerations for their definitions.

Multiple entries in propositional memory: The question whether the same surface form of a word has multiple representations in semantic or propositional memory has been investigated in a number of different processing tasks. Usually, this issue has centered around the difference between homographs (words with two clearly different meanings, such as "bank") and nonhomographs. However, as indicated below, homography is probably best viewed as a continuous rather than an all-or-none characteristic of words.

The multiple representation question has taken two forms. The first is whether or not multiple entries for homographs exist, and the answer to this seems to be a clear Yes. The second is whether subjects can be simultaneously aware of more than one of the meanings of a homograph, and the answer to this seems to be a clear No, with however, an interesting exception.

Both questions have been investigated using the lexical decision task, in which subjects must decide whether a string of letters is a word or a nonword. Several investigators have found that homographs are responded to faster than nonhomographs (Rubenstein, Garfield, & Milliken, 1970; Rubenstein, Lewis, & Rubenstein, 1971; Jastrzembski & Stanners, 1975). Furthermore, semantic priming of a word, by presenting a related word prior to the target word, speeds word classification (Meyer & Schvaneveldt, 1971; Schvaneveldt & Meyer, 1973). These results, taken together, suggest that semantic activation of one node in semantic memory spreads to other, semantically related nodes. This is known as the spreading-activation hypothesis.

Exactly the same paradigm has been used to investigate the second question: whether subjects can be aware of more than one meaning of a word at a time. Schvaneveldt, Meyer, and Becker (1976) found that when one meaning of a homograph

was primed, it facilitated lexical decisions only for words related in meaning to the primed meaning, not to words related to the unprimed meaning (e.g., priming "bank" for its "river" sense speeds the classification of river, but not of "money", and vice versa). This suggests that only one meaning of a word can be active in semantic memory at any one time.

In an interesting variation on the preceding paradigm, Marcel (1979) used a pattern mask to make the homograph unavailable to consciousness. With this procedure, he found facilitation for words related to both primed and unprimed meanings of the homograph. This suggests that when propositional memory is activated at an unconscious level, all meanings and their related concepts are activated.

The concept of multiple access and representation in semantic memory has also been used as a principle to account for the encoding specificity results obtained by Tulving and his associates (Tulving & Thomson, 1971). The encoding specificity results show that changing the context in which a word is imbedded between study and test produces a decrement in recognition memory over no change in context. Tulving argues from these results that (a) recognition, contrary to popular belief, does involve a retrieval operation; and (b) that an episodic memory store is established, during episodic memory tasks, separately from the semantic store.

Critics of the encoding specificity principle argue that most words are homographic to some degree, even though they might not be classified as such with an all-or-none criterion. Therefore the node accessed in semantic memory for a word occurring in one context will be different from the node accessed when the same surface form occurs in a second context, thereby accounting for the decrement in recognition memory with changed-context tests (Martin, 1975; Anderson & Bower, 1974; Reder, Anderson & Bjork, 1974).

The principle of multiple access is represented in the model by assuming that a given entry in the acoustic image store may have more than one entry in propositional memory (e.g., "top", "well", "bank"), and also, when picturable, more than one entry in the visual image store. When the particular meaning of a word is disambiguated (either spontaneously by the subject or by experimenter-imposed context), this results in activation of a single node in propositional memory whose output in turn may activate a single entry in the visual image store.

Dual-coding versus propositional memory models: The dual-code model, most widely associated with the work of Paivio and his students (Paivio, 1969, 1971; Paivio & Csapo, 1973) proposes that two qualitatively different memory stores exist -- one, the visual image system, in which information about picture or object appearance is stored in an analogue form, having spatial and configural properties similar to the picture (or its object), and the second, verbal system, in which both phonemic/acoustic and symbolic information are stored. The two systems differ in neural locus and in their specialization for different tasks, but are richly interconnected. In Paivio's system, meaning is provided by associations between and within each store.

Both pictures and words may be encoded into both systems. The likelihood is high that pictures will be encoded into the image systems and words into the verbal system, pictures are less likely to be encoded into the verbal system and concrete words even less likely to be encoded into the image system. Since the two memory systems are assumed to be independent, registration in either one or both of the stores may provide a basis for recognition or recall.

Pictures are better remembered than words in most tasks for one or both of two reasons: first, pictures are more likely to be dually-coded than words, i.e., registered in both the image and verbal stores; and second the image code, which is more likely to be stored to a picture than to a word, is the more effective code (Paivio & Csapo, 1973).

Evidence that naming of pictures is more likely than imaging of concrete words was obtained by Snodgrass and McClure (1975) in an item recognition task, and evidence that the image code is more effective than the verbal code was obtained by Paivio and Csapo (1973) in a recall task, and Snodgrass, Burns, and Pirone (1978) in an item and order recognition task. However, Paivio found that the image code was equally effective regardless whether it was generated to a picture or to a word when recall was the memory test, whereas Snodgrass et al. found that the image code was more effective when generated to a picture than to a word when recognition was the memory test. We shall return to this difference in a moment.

A second model which accounts for the superiority of pictures over words in episodic memory is the superior sensory code hypothesis, proposed by Nelson and his colleagues (Nelson, Reed, & McEvoy, 1977; Nelson, Reed, & Walling, 1976). Their model is similar to the present one in proposing that both pictures and words access a common semantic representation. However, their model differs from this one in proposing that the sole advantage of pictures over words is the more elaborate sensory codes of pictures as compared to words. They showed that when the visual similarity of a set of pictures which served as stimulus items in a paired-associate recall task was increased, pictures showed no advantage over words as stimulus items, and in some cases a word advantage was obtained.

The most persuasive critics of the dual-code hypothesis have been the propositional theorists (Anderson & Bower, 1973; Pylyshyn, 1973, 1978) who argue that both visual and verbal information are encoded in a common format, as a system of abstract propositions and relationships. The propositionalists do not deny the introspective reality of mental imagery or of verbal thinking, but simply regard them as epiphenomena which do not in themselves constitute evidence for two types of memory storage.

The present model includes aspects of both the dual-coding and propositional models. The dual-coding system is represented by the two image systems at Level II. In contrast to Paivio's model, however, interconnections between the two systems are typically made via the propositional system, rather than directly. Comparisons among entries in the two image systems are made only on the basis of shared feature values: thus, phonemic and visual similarity judgments can be made within Level II, but conceptual similiarity judgments must be made by accessing Level III.

The propositional level of the proposed model corresponds, of course, to the propositional memory of the propositional theorists, including the assumption made by Pylyshyn (1973) that the operation of the propositional system is not available to consciousness.

Relationship of empirical data to the model

In this review of empirical evidence for processing differences between pictures and words, I shall concentrate on two issues: first, how pictures and words differ in their access to semantic memory, and second, how they differ in their registration in episodic memory.

Evidence from semantic memory tasks.

Semantic memory tasks may be divided into those which require translation from one internal image code to the other, such as naming pictures and imaging words, and those which require access to propositional information such as categorization and symbolic comparison judgments.

Naming and imaging compared: In the model, naming a word or picture involved accessing the acoustic image store, and imaging a picture or word involves accessing the visual image store.

Naming latencies, as measured by voice key activation, are faster for words than for pictures (Cattell, 1886; Fraisse, 1960; Potter & Faulconer, 1975). Each of the following variables has been shown to affect picture-naming latencies; the frequency in print of a picure's name (Oldfield & Wingfield, 1965); age of

acquisition of the concept name (Carroll & White, 1973); and the codability, or degree of naming agreement, of the picture, as measured by the information statistic H (Lachman, 1973).

Although direct comparisons between naming and imaging latencies are hard to make because different responses have been used, available evidence suggests that imaging a word takes longer than naming a picture. Imaging latencies have been measured in single key-press tasks, in which subjects are to press a key as soon as the image has been generated (e.g., Paivio, 1966; Kosslyn, 1975) or in two-choice tasks, in which subjects are to press a key as soon as they have identified as present or absent a particular aspect of an image (Kosslyn, 1973; 1978). Single keypress latencies have proved sensitive to variables related to the image. For example, Paivio (1966) found that images could be generated to concrete words faster than to abstract words, and Kosslyn (1978) found that forming images of complex pictures took longer than those for simple pictures of the same concepts. Furthermore, identifying a detail in a small image took longer than in a large image (Kosslyn, 1975).

In a two-key response situation, Kosslyn and his collaborators have found that scanning a larger distance in an image takes more time than scanning a small distance (Kosslyn, Ball & Reiser, 1978). In addition, Kosslyn and Shwartz (1977) have implemented a computer program of image generation which shows many of the properties their human subjects show. Their program is similar to the proposed model in that potential (stored) image values are distinguished from generated images (formed on a cathode ray tube) and that default values are given for certain values such as size. Their program differs in having both a "literal representation," a rough drawing whose visual features are stored in a spatial grid, and propositional features which

can be called to elaborate and enrich the image.

Within the model, naming words is faster than naming pictures because direct access between physical analysis and the acoustic image code is possible for words whereas pictures must first access their visual image code and then either find direct access to the acoustic image code or access it via the propositional system.

The variables of frequency and age-of-acquisition have much larger effects on picture-naming latencies than on word-naming latencies. This suggests that the effect may take place at the propositional level: since pictures are more likely to require propositional access than words, they will show correspondingly larger effects. In contrast, picture codability would appear to have its effect in the access route from the propositional to the acoustic image code. If a picture has more than one name, a particular propositional node may access more than one acoustic image code. These multiple representations then might be expected to slow naming for pictures but not for their names.

That imaging words takes longer than naming pictures can be accounted for in at least two ways. First, since words are semantically ambiguous, nodes may be accessed which have no corresponding visual image code (e.g., the adverbial form of "well") or several which compete (e.g., the various possible images of "bat" and "top"). Second, since visual images have more variable feature values than acoustic images, visual images might take longer to generate than acoustic images.

Categorizing names versus pictures: It seems clear that words can be named faster than pictures, and that pictures can be imaged (that is, their central or defining physical features extracted) faster than words. What

can we say about the speed with which pictures and words can be understood? While understanding the meaning of pictures and words can take various forms (e.g., the symbolic comparison judgment, Banks, 1978), this review will concentrate on the categorization task.

The available evidence on categorization times suggests that pictures are categorized more quickly than words. For example, in a Yes/No task in which the category is given ahead of time, pictures are either categorized faster than words (Potter & Faulconer, 1975) or with the same speed (Smith & Magee, 1979). In a same-different task, when two instances of pictures or words in the same or different category are presented simultaneously, pictures result in faster match or mismatch decisions than words (Pellegrino, Rosinski, Chiesi & Siegel, 1977). Furthermore, in a Yes/No categorization task in which pictures and words are presented simultaneously, incongruous pictures have a larger interference effect on word categorization than incongruous words have on picture categorization (Smith & Magee, 1979).

Characteristics of the category exemplars, and the similarity of the items within categories, have also been shown to affect categorization times. Rosch (1975) showed that prototypical members were categorized faster than nonprototypical members, and Klatsky and Ryan (1978) and Klatsky and Stoy (1974) showed that the physical similarity of items within a category affected both category matches and identity matches.

Within the context of the model, categorization superiority of pictures over words is puzzling. A possible basis for it might be, again, the greater semantic ambiguity of words over pictures. However, because the context provided by the categorization task would appear to bias the meaning of words that are homographic toward their intended meaning, and since the evidence suggests that only one meaning of a word is activated in propositional memory, this is a tenuous explanation at best.

There are at least two other possible alternatives. One is that objects and their pictorial representations within a category may appear more similar visually than those between categories, whereas the objects' names would show no such graphemic similarity. Thus, some judgments might be made on the basis of feature similarity (i.e., at Level II) so that propositional memory need not be accessed. This interpretation is consistent with the results found by Klatsky and her collaborators for artificially defined categories.

Alternatively, it may be that subjects in a categorization task verbally rehearse the name of the category when they are carrying out the Yes/No or matching task. This results in verbal interference with the naming of words (required before they can access semantic memory) but not the recognition (imaging) of pictures.

Evidence from episodic memory tasks.

The model makes no distinctions between episodic and semantic memory. Rather, it assumes that the presentation of a to-be-remembered item in the study phase of an episodic memory experiment may cause that item to be registered in three stores: the mismatch accumulator store which registers the degree and type of discrepancy between the physical input and the prototypical image; the prototypical image store; and the propositional store.

Consistent with the levels-of-processing notion (Craik & Lockart, 1972; Craik & Tulving, 1975), I assume the registration to be hierarchical, with registration in the deepest, propositional store dependent upon registration in the shallower stores. I also assume that the rate at which registration information decays over time is fastest for the shallowest levels and slowest for the deepest level. Thus, the typeface or voice in which a word is presented is available for a shorter amount of time than whether a concept was presented

as a picture or a word, which in turn is available for a shorter amount of time than the meaning of the concept.

During the memorization phase of an episodic memory task, subjects may store information at all three levels of the memory system. The importance of each level's contribution to memory performance will then depend upon at least two variables: first, the delay between study and test, and second, the type of memory test.

I shall consider results from three types of test procedures: same-form recognition, same-concept recognition, and recall. In same-form recognition, recognition memory for a picture is assessed by presenting the same picture, and recognition memory for a word is assessed by presenting the same word. In same-concept recognition, recognition memory for a picture is assessed by presenting the picture's name, and recognition memory for a name is assessed by presenting the name's picture.

It seems plausible that same-form recognition tests should be more effective in accessing all three levels of the memory system than opposite-form recognition tests, but that, to the extent that information is still available at the two shallower levels, subjects in opposite-form conditions should be able to decide, with fair accuracy, in which form the original concept was presented.

Studies which have compared performance on same-form versus opposite-form recognition have shown that same-form performance is either superior to or equal to opposite-form performance. For example, Jenkins, Neale, and Deno (1967) found a decrement with opposite-form over same-form recognition tests, whereas Paivio (1976), Snodgrass and McClure (1975), and Snodgrass and Asiaghi (1977) found no such decrement. In addition, Snodgrass and McClure (1975) found that memory for the form in which a concept was studied was remarkably good, even when the concept was tested in its opposite form. In fact, they found that when a signal detection model was used to analyze memory for both concept and form, there was no significant difference between the d' estimates for concept memory and for form memory.

These results appear, at first glance, to cast some doubt on the present model's assumption that registration occurs at all three levels. If this were the case, subjects should have been more uncertain about which form the original stimulus was in (form memory), than which concept had been presented (concept memory), because form memory is only retrievable from the shallower levels, whereas concept memory is retrievable from all three levels.

However, two cautionary remarks need to be made: First, subjects in this experiment were warned prior to the study phase that recognition tests would be in either form, and thus they knew they would need to remember both the surface and propositional form of the concept. Secondly, the picture stimuli that were used consisted of some inherently ambiguous stimuli, so all picture-name pairs were presented to the subjects prior to the experiment. This may have provided subjects with ways in which to distinguish images stored to pictures from those stored to words.[1]

Recall, relying as heavily as it does on retrieval operations, may differ from recognition in the importance which the various levels have in determining performance. Available evidence suggests that the meaning, rather than the appearance, of concepts is more important in recall than in recognition, and thus that registration at the propositional level plays a much larger role than registration at the shallower levels.

For example, Frost (1972) showed that when pictorial stimuli varied on both semantic similarity (category membership) and visual similarity (shape and orientation of the drawing), subjects expecting a recall task clustered by semantic category in their (expected) recall test, whereas subjects expecting a recognition test clustered by both dimensions in their (unexpected) recall test. Furthermore, there is a large verbal learning literature which shows

that semantic organizational factors play a much larger role in recall than in recognition, where their effect is either absent or minimal (see Klatzky, 1975 for a review of this evidence). This evidence leads to two conclusions: first, that recall is much more dependent upon retrieval from propositional memory than from the shallower stores based on appearance, and second, that subjects can vary their encoding strategies to take advantage of this fact.

This view is consistent with the recall and recognition differences between dually-encoded pictures and words reviewed earlier. According to the model, dually-encoded pictures and words are assumed to be registered in both image stores, as well as in propositional memory. In addition, dually-encoded pictures have an advantage over dually-encoded words in accumulating more mismatch information. However, this mismatch information may only be effective in a recognition test, and not in recall. This would account for the finding that imaged words do not differ from pictures in recall (Paivio & Csapo, 1973) yet imaged words are inferior to pictures in recognition (Snodgrass & McClure, 1975; Snodgrass, Burns, & Pirone 1977).

If a picture and its name both access the same propositional node, it should be possible to show some confusion between the two surface forms. Just such an effect was observed by Snodgrass, Wasser, Finkelstein, and Goldberg (1974), who found that when subjects were forced to choose between two surface representations of the same concept in a two alternative forced-choice recognition test, they showed more confusions than would be expected from control conditions. These confusions were particularly large for unambiguously nameable studied pictures paired with their names, and were nonexistent for studied words paired with their pictures. The lack of a confusion effect for words was attributed either to the absence of stored images for words, or, if images had been stored, the greater variability of image codes, leading to the greater liklihood of a mismatch between the generated image to a name and that name's picture.

This has been a selective review of the episodic memory literature on picture-word differences. The results are understandable, if not predictable, within the context of the proposed model. It is clear that more sensitive tests of the model will have to be devised before either accepting it, modifying it, or rejecting it.

Footnote

[1]This points out the desirability of having a large set of picture stimuli available which do not suffer from either name or image ambiguity. Such a set has recently been published (Snodgrass and Vanderwart, 1979).

References

Anderson, J.R., & Bower, G.H. Human associative memory. Washington, D.C.: V.H. Winston, 1973.

Anderson, J.R., & Bower, G.H. A propositional theory of recognition memory. Memory & Cognition, 1974, 2, 406-412.

Banks, W.P. Encoding and processing of symbolic information in comparative judgments. In G.H. Bower (Ed.), The psychology of learning and motivation (Vol 2). New York: Academic Press, 1977.

Carroll, J.B., & White, M.N. Word frequency and age of acquisition as determiners of picture-naming latency. Quarterly Journal of Experimental Psychology, 1973, 25, 85-95.

Cattell, J.M. The time it takes to see and name objects. Mind, 1886, 11, 63-65.

Collins, A.M., & Quillian, M.R. Experiments on semantic memory and language comprehension. In L.W. Gregg (Ed.), Cognition in learning and memory. New York: John Wiley, 1972. (a)

Collins, A.M., & Quillian, M.R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972. (b)

Collins, A.M., & Loftus, E.F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Craik, F.I.M., & Lockart, R.S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Craik, F.I.M., & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104, 268-294.

Fraisse, P. Recognition time measured by verbal reaction to figures and words. Perceptual & Motor Skills, 1960, 11, 204.

Frost, N. Encoding and retrieval in visual memory tasks. Journal of Experimental Psychology, 1972, 95, 317-326.

Jastrzembski, J.E., & Stanners, R.F. Multiple word meanings and lexical search speed. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 534-537.

Jenkins, J.R., Neale, D.C., & Deno, S.L. Differential memory for picture and word stimuli. Journal of Educational Psychology, 1967, 58, 303-307.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Klatzky, R.L. Human memory: Structures and processes. San Francisco: Freeman, 1975.

Klatzky, R.L., & Ryan, A.S. Category-structure effects in picture comparisons. Perception & Psychophysics, 1978, 23, 193-204.

Klatzky, R.L., & Stoy, A.M. Using visual codes for comparisons of pictures. Memory & Cognition, 1974, 2, 727-736.

Kosslyn, S.M. Scanning visual images: Some structual implications. Perception & Psychophysics, 1973, 14, 90-94.

Kosslyn, S.M. Information representation in visual images. Cognitive Psychology, 1975, 7, 341-370.

Kosslyn, S.M. Imagery and internal representation. In E. Rosch and B.B. Lloyd (Eds.), Cognition and categorization. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Kosslyn, S.M., Ball, T.M., & Reiser, B.J. Visual images preserve metric spatial information: Evidence from studies of image scanning. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 47-60.

Kosslyn, S.M., & Shwartz, S.P. A data-driven simulation of visual imagery. Cognitive Science, 1977, 1, 265-296.

Lachman, R. Uncertainty effects on time to access the internal lexicon. Journal of Experimental Psychology, 1973, 99, 199-208.

Marcel, A.J. Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In R.S. Nickerson (Ed.), Attention and Performance VIII. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

Martin, E. Generation-recognition theory and the encoding specificity principle. Psychological Review, 1975, 82, 150-153.

Meyer, D.E., & Schvaneveldt, R.W. Facilitation in recognizing words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 1971, 90, 227-234.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

Nelson, D.L. Remembering pictures and words: Appearance, significance, and name. In F.I.M. Craik & L. Cermak (Eds.), Levels of Processing. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

Nelson, D.L., Reed, V.S., & McEvoy, C.L. Learning to order pictures and words: A model of sensory and semantic encoding. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 485-497.

Nelson, D.L., Reed, V.S., & Walling, J.R. Pictorial superiority effect. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 523-528.

Norman, D.A., & Rumelhart, D.E. Explorations in cognition. San Francisco: Freeman, 1975.

Oldfield, R.C., & Wingfield, A. Response latencies in naming objects. Quarterly Journal of Experimental Psychology, 1965, 17, 273-281.

Paivio, A. Latency of verbal associations and imagery to noun stimuli as a function of abstractness and generality. Canadian Journal of Psychology, 1966, 20, 378-387.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart, & Winston, 1971.

Paivio, A. Imagery in recall and recognition. In J. Brown (Ed.), Recall and recognition. New York: Wiley, 1976.

Paivio, A., & Csapo, K. Picture superiority in free recall: Imagery or dual coding? Cognitive Psychology, 1973, 5, 176-206.

Pellegrino, J.W., Rosinki, R.R., Chiesi, H.L., & Siegel, A. Picture-word differences in decision latency: An analysis of single and dual memory models. Memory & Cognition, 1977, 5, 383-396.

Potter, M.C., & Faulconer, B.A. Time to understand pictures and words. Nature, 1975, 253, 437-438.

Pylyshyn, Z.W. What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin, 1973, 80, 1-24.

Pylyshyn, Z.W. Imagery and artificial intelligence. In W. Savage (Ed.), Perception and cognition: Issues in the foundations of psychology. Minneapolis, Minn.: University of Minnesota Press, 1978.

Quillian, M.R. Word concepts: A theory and simulation of some basic semantic capabilities. Behavioral Science, 1967, 12, 410-430.

Reder, L.M., Anderson, J.R., & Bjork, R.A. A semantic interpretation of encoding specificity. Journal of Experimental Psychology, 1974, 102, 648-656.

Rips, L.J., Shoben, E.J., & Smith, E.E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 1-20.

Rosch, E. Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 1975, 104, 192-233.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. Basic objects in natural categories. Cognitive Psychology, 1976, 8, 382-439.

Rosch, E. Principles of categorization. In E. Rosch and B.B. Lloyd (Eds.), Cognition and categorization. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Rubenstein, H., Garfield, L., & Milliken, J.A. Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 487-494.

Rubenstein, H., Lewis, S.S., & Rubenstein, M.A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Schvaneveldt, R.W., & Meyer, D.E. Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, 1973.

Schvaneveldt, R.W., Meyer, D.E., & Becker, C.A. Lexical ambiguity, semantic context, and visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 243-256.

Smith, E.E. Theories of semantic memory. In W.K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 5). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Smith, E.E., Shoben, E.K., & Rips, L.J. Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 1974, 81, 214-241.

Smith, M.C., & Magee, L.E. Tracing the time course of picture-word processing. Journal of Experimental Psychology: General, 1979 (In Press).

Snodgrass, J.G., & Asiaghi, A. The pictorial superiority effect in recognition memory. Bulletin of the Psychonomic Society, 1977, 10, 1-4.

Snodgrass, J.G., Burns, P.M., & Pirone, G.V. Pictures and words and space and time: In search of the elusive interaction. Journal of Experimental Psychology: General, 1978, 107, 206-230.

Snodgrass, J.G., & McClure, P. Storage and retrieval properties of dual codes for pictures and words in recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 521-529.

Snodgrass, J.G., & Vanderwart, M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory, 1979 (In Press).

Snodgrass, J.G., Wasser, B., Finkelstein, M., & Goldberg, L.B. On the fate of visual and verbal memory codes for pictures and words: Evidence for a dual coding mechanism in recognition memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 27-37.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Tulving, E., & Thomson, D.M. Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 1971, 87, 116-124.

Wicklegren, W.A. Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. Psychological Review, 1979, 86, 44-60.

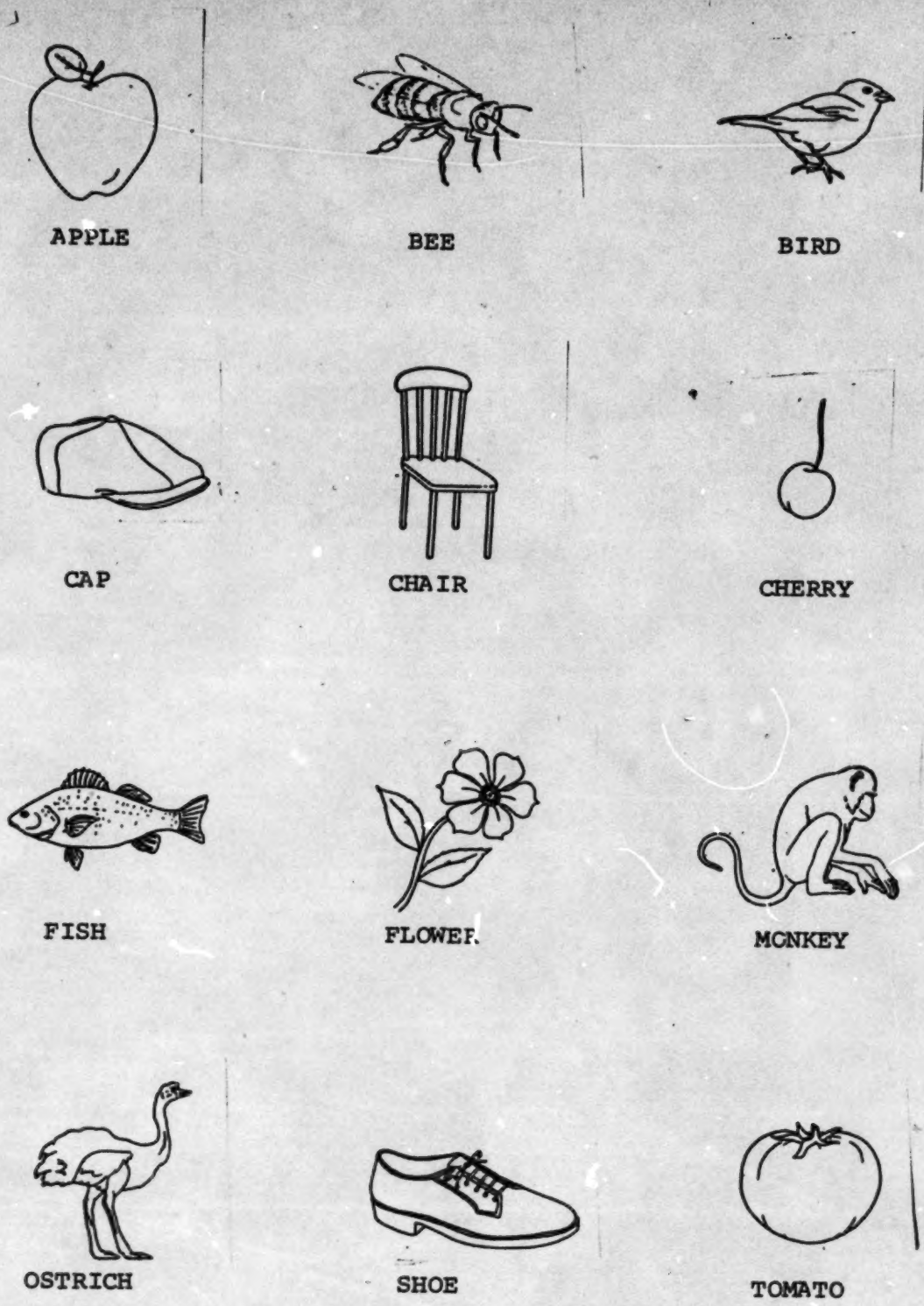

Figure 1. Exemplary picture-word pairs considered by the proposed model. (from Snodgrass & Vanderwart, 1979).

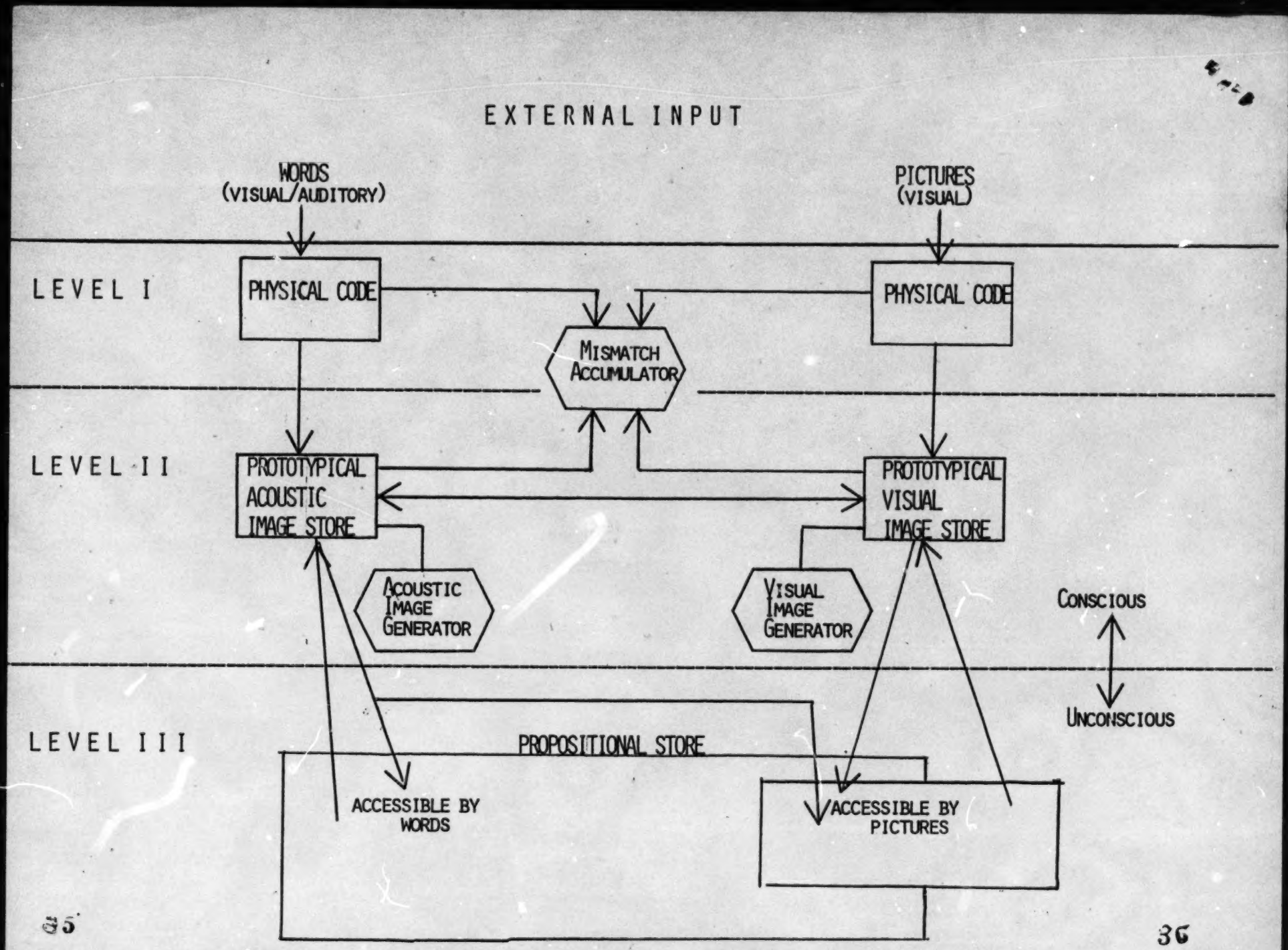

Figure 2. A schematic diagram of a model for picture and word processing.